

Line graphs as social networks

M. J. Krawczyk^{1a}, L. Muchnik^{2b}, A. Mańka-Krasoń^{1c}
and K. Kułakowski^{1d}

¹*Faculty of Physics and Applied Computer Science, AGH University of
Science and Technology, al. Mickiewicza 30, PL-30059 Kraków, Poland*

²*Information, Operations and Management Systems Department, Stern School
of Business at the New York University, 44 West Fourth Street New York, NY
10012, USA*

^agos@fatcat.ftj.agh.edu.pl, ^blevmuchnik@gmail.com,
^cimpresja@gmail.com, ^dkulakowski@novell.ftj.agh.edu.pl

October 13, 2010

Abstract

The line graphs are clustered and assortative. They share these topological features with some social networks. We argue that this similarity reveals the cliquey character of the social networks. In the model proposed here, a social network is the line graph of an initial network of families, communities, interest groups, school classes and small companies. These groups play the role of nodes, and individuals are represented by links between these nodes. The picture is supported by the data on the LiveJournal network of about 8×10^6 people. In particular, sharp maxima of the observed data of the degree dependence of the clustering coefficient $C(k)$ are associated with cliques in the social network.

PACS numbers:

Keywords: social networks, scale-free networks

1 Introduction

In mathematically oriented sociology, a social network is a paradigm. The idea is to express social relations between individuals by weighted or unweighted links between nodes of a graph. Rough as it is, this approximative representation got a wide interest of social scientists [1, 2, 3, 4, 5]. In interdisciplinary areas, the research on networks was boosted by the seminal paper of Watts and Strogatz in 1998 [6]. Since then, several books are published on various applications of networks [7, 8, 9, 10, 11]. Our aim here is to foster a new application of the network formalism: we propose that the structure of some social networks makes them similar to the structure of the line graph, constructed on a scale-free growing network. The line graphs are known for at least 80 years [12, 13], but in the above mentioned interdisciplinary stream their relevance seems underestimated. Some recent applications of this kind of graphs can be found in [13, 14, 15, 16, 17, 18, 19].

Our argumentation can be sketched as follows. Suppose that the actual structure of some society is a set of cliques or almost fully connected clusters which can be identified with families, groups of friends, small companies or interest groups. Each such clique can be represented as a node of an otherwise uncorrelated network. As it was indicated in a recent analysis of information

spreading, two small and strongly connected groups are linked by individuals who belong to both of them and contribute in the information transfer between them [20]. Once we are interested in a construction of a conventional social network, where humans play the role of nodes, the network should be constructed as the line graph of the initial network of families, school classes etc.

A fact rarely considered by social network researchers is that social ties represent particular social context such as common interest, association with the same group or collocation. Considering this, the natural organization of humans in nearly fully connected groups linked by their simultaneous participation in few of them leads to emergence of social graph observed by researchers. In this manuscript we show that a line graph transformation applied to the underlined network of such groups may indeed lead to a graph with properties commonly observed in social networks.

In the next section we summarize the arguments of Newman and Park, that social networks are both transitive (clusterized) and assortative [21]. In Section 3 we refer to our recent calculations on the line graphs, where we have shown that these graphs are both clusterized and assortative [22, 23]. In Section 4 we describe new data on the network of users of LiveJournal, which also support the above characteristics of social networks. In the same section the data plot is shown on the clustering coefficient C as dependent on the node degree k . In Section 5 we compare the plot $C(k)$ with the result of simulations on the line graph, formed from an uncorrelated scale-free network. Last section is devoted to conclusions.

2 Social networks

In [21] and literature therein ([24, 25, 26, 27, 28] among others) Newman and Park bring examples of social networks which are clusterized and assortative. Let us recall that in unweighted networks, the clustering is measured by the clustering coefficient C defined as

$$C = \sum_i \frac{2y_i}{k_i(k_i - 1)} \quad (1)$$

where y_i is the actual number of links between neighbours of i -th node, and k_i is the degree of this node, i.e. the number of its neighbours. In other words, the clustering coefficient is the probability that two neighbours of a node are mutually connected. Once a node has zero or one neighbour only, its contribution to C is zero. By clusterized we mean that the clustering coefficient C is clearly larger, than for a random network. The examples are: the network of film-actor collaborations ($C=0.20$), the collaboration network of mathematicians ($C=0.15$), the network of company directors ($C=0.59$) and an e-mail network ($C=0.17$). In all these examples, the respective values of C for non-clusterized counterparts are smaller by at least one order of magnitude.

The assortativity is a tendency of highly connected nodes to have highly connected neighbours. It can be measured by the Pearson correlation coefficient r

between degrees of neighboring nodes. Once r is positive for some network, we can classify this network as assortative. Indeed, as demonstrated by Newman in [25], the coefficient r is positive for the network of film-actor collaborations ($r=0.208$), the coauthorship networks of mathematicians ($r=0.12$), physicists ($r=0.363$) and biologists ($r=0.127$), the network of company directors ($r=0.276$) and an e-mail network ($r=0.17$). An exception was found for the network of romantic (not necessary sexual) relationships between students at a US high school [29], where $r=-0.029$. However, in this particular case we can understand that individuals are not willing to involve third part into their romantic relations. An alternative way to check if a network is assortative or not is to plot the mean degree k' of neighbours of nodes of degree k as dependent on k . If the plot $k'(k)$ is ascending, the investigated network is assortative: more connected nodes have on average more connected neighbours. As we checked in [23], for artificially generated uncorrelated scale-free networks k' does not depend on k .

A widely cited explanation for emergence of assortativeness and high clustering in social network [21] is that individuals belong to many groups, and their social connections are limited to members of the groups they belong to. In other words, the social structure is equivalent to a bipartite network of groups and individuals. In this network, individuals are connected to groups they belong to. What is usually observed is a projection of this structure onto just the individuals. As an outcome of this projection, we have a social network where individuals are connected with probability p if they belong to the same group; otherwise they are not connected. This model network is found to be both clusterized and assortative [21].

We suggest that applying the line graph transformation to a network of groups such as families, school classes and common interests associations representing cliques of individuals and connected by the same person that simultaneously belongs to several groups, can result in a social network closely resembling the ones we actually observe.

This quantitative explanation of the clusterized and assortative structure of social networks agrees with the well-established opinion of social scientists that cohesive small groups are the main motif in a society. In [3], a list is given of four general properties of these groups. These are: the mutuality of ties, the closeness or reachability of group members, the frequency of ties among members and the relatively smaller frequency of the ties among non-members. In the network formalism, all these properties find their formal shape. On the other hand, the quantitative search of communities initiated in [26] developed into a large branch of science of networks [11]; recent review on this search can be found in [30].

3 Line graphs

From a given graph G of N nodes and L links, a line graph G' can be constructed as follows [31]. A node of G' is assigned to each link of G . Two nodes of G' are linked if and only if the respective links in G shared a node. In this way, the number N' of nodes in G' is equal to the number L of links in G . The

number L' of links in G' depends on the degree distribution $P(k)$ in G . We have shown numerically in [22], that for three kinds of networks (Erdős-Renyi networks, the growing exponential networks and the growing scale-free networks) the degree distribution of G' is close to the degree distribution of G . Basically, a node of degree k is converted to a clique (fully connected graph) of k nodes and $k(k-1)/2$ links. Further, a link in G joining nodes of degrees k_1 and k_2 is converted into a node in G' of degree $k_1 + k_2 - 2$ which belongs to two cliques, one of k_1 nodes and another of k_2 nodes.

For an uncorrelated graph G of the mean degree $\langle k \rangle$ much smaller than N we can assume, that two different neighbours of a node are not mutually linked. A contribution to the clustering coefficient C of a node in G' contains then merely the contributions from two separate cliques. The number of links between $k_1 + k_2 - 2$ neighbours is then $(k_1 - 1)(k_1 - 2)/2 + (k_2 - 1)(k_2 - 2)/2$. For the degree distribution $P(k)$, the clustering coefficient C is

$$C = \frac{\sum_{k_1, k_2} k_1 P(k_1) k_2 P(k_2) \frac{(k_1-1)(k_1-2) + (k_2-1)(k_2-2)}{(k_1+k_2-2)(k_1+k_2-3)}}{\sum_{k_1, k_2} k_1 P(k_1) k_2 P(k_2)} \quad (2)$$

This formula was used in [22] and the results were compared with numerical calculations. Both methods confirm that for $\langle k \rangle$ greater than 5, the clustering coefficient C is not smaller than 0.5.

The same methods were applied to demonstrate that the line graphs constructed on uncorrelated networks are assortative [23]. This is a direct consequence of the fact that the neighboring nodes in the line graphs are formed from links sharing a common node in the initial graph. The degree of this common node contributes to the degree of both neighboring nodes in the line graph.

4 LiveJournal

LiveJournal [32] is a remarkably popular platform for personal blog management, populated with over 8 million blogs and over 1 million of communities. LiveJournal was among the first of such platforms available online and it still remains one of the most active and popular. Its users manage personal blogs where they share their daily experiences, political views or discuss news events. Users can also comment on posts of other users.

Unlike more dynamic systems like Facebook and Twitter that gained their popularity rather recently, LiveJournal is not based on personal messages or applications. Typical LiveJournal post may contain a significant amount of text with embedded images or video and may be followed by discussion that in times exceed thousands of comments.

The LiveJournal system encourages users to bookmark and monitor particular blogs. This feature is exercised by virtually all users and results in a network of references between these blogs. The vast majority of blogs regularly read by a person are typically stored in the form of bookmarks as part of his profile. This degree of penetration of this behavior is driven by two main reasons. First,

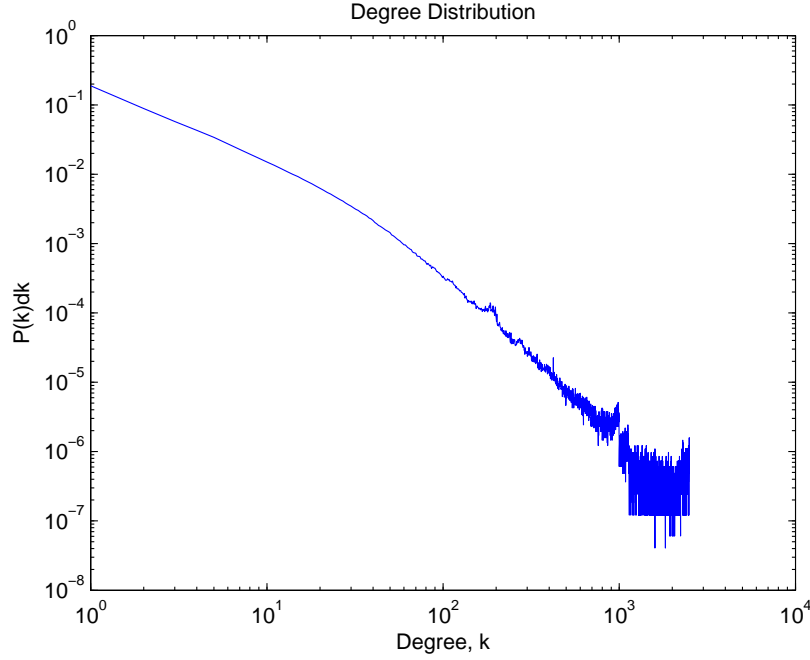


Figure 1: The degree distribution of the social network of LiveJournal.

its convenience: it is impractical to periodically search for a particular blog and check whether it has new posts. The system automatically notifies users of the updates to the bookmarked blogs. Second, to protect their privacy, many users limit visibility of their posts to the users listed in their list of friends. Overall, the personal nature of these blogs and the intimate relationship between their authors give this network a powerful social aspect. In fact, we conducted a large number of case studies analyzing the threads of comments to verify that authors of many of the connected blogs actually know each other in person. It is therefore legitimate to refer to the network of blog bookmarks as social network.

In addition to personal profiles, users create communities that are in fact blogs run in collaboration by a number of users. Communities usually spin around a particular interest, well defined topic or represent a group of people united by a common task (such as, for instance, role-playing games) but otherwise are very similar to personal blogs. Periodic posts are discussed in threads of comments.

LiveJournal has been used in a large number of academic studies [33, 34, 35, 36, 37, 38, 39] due to its openness and availability of its well-designed APIs [<http://www.livejournal.com/developer/>]. In particular, all user profiles including the lists of monitored personal blogs and communities along with detailed information about the blog owners and their interests are freely accessible.

The data used in this work was obtained by crawling LiveJournal and col-

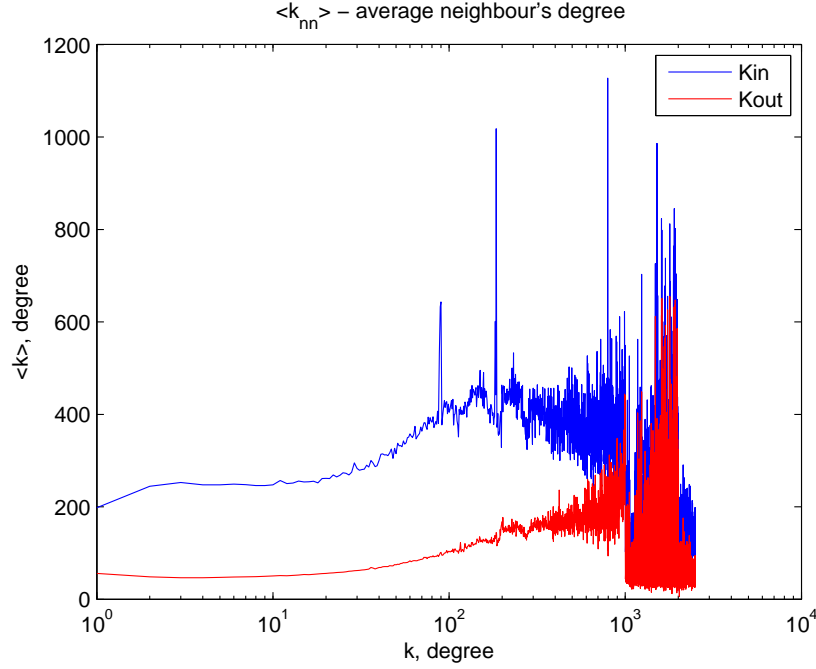


Figure 2: Mean in-degree (blue) and out-degree (red) of neighbours of nodes of degree k for LiveJournal.

lecting the entire content of all user profiles in the giant component. We defined the network nodes to correspond to personal blogs. Directional links connecting these nodes represent the record that a particular user (owning one blog) monitors another blog (owned by another user). We disregard community blogs as they usually do not represent individual users (but rather a group) and cannot be considered as part of the social network represented by personal blogs.

The social network obtained from the crawl contains 8.1 million users and over 125 million links. The average clustering coefficient is $C = 0.1522$. However, having excluded nodes of degree 0 and 1, we get $C = 0.2684$. The degree distribution is shown in Fig. 1. The log-log plot reveals some deviations from linearity; we can distinguish two ranges of k , between 1 and 50 and between 50 and 1000, where it is linear. Still, the slopes of the curves within these ranges do not differ much. The assortativity is confirmed by the data shown in Fig. 2. There, the plots $\langle k'(k) \rangle$ increase with k both for in-going and out-going links.

As shown in Fig. 3, the clustering coefficient $C(k)$ of LiveJournal varies strongly with k in the social network. A strong local peak of the function $C(k)$ at given k means that there is a lot of nodes of almost the same degree which are strongly clusterized. It is straightforward to imagine that these nodes belong to the same cluster - a clique. As cliques are formed from nodes in line graphs, the latter should display also the same kind of oscillations of the function $C(k)$. Indeed we found this behaviour for an artificially generated line graph of 10^4

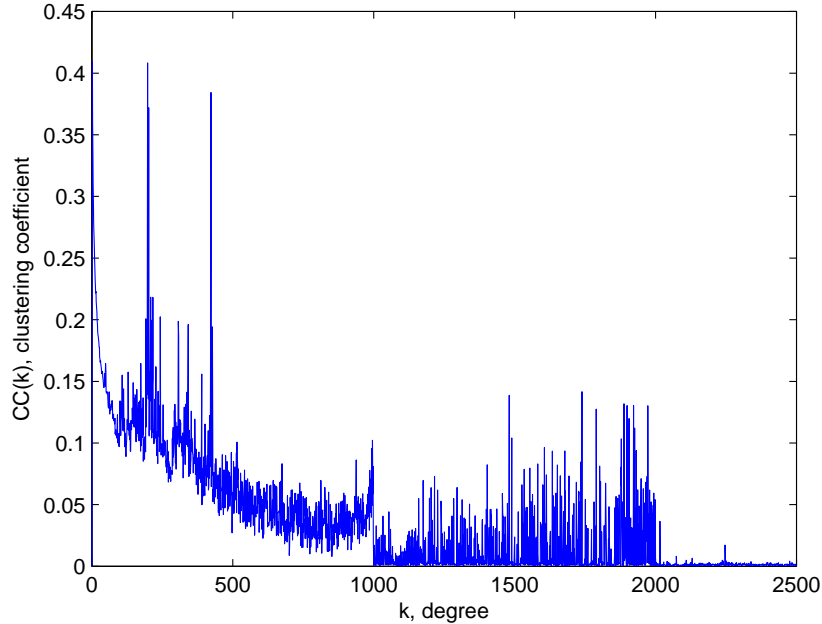


Figure 3: The clustering coefficient $C(k)$ against the node degree for LiveJournal.

nodes, as shown in Fig. 4. However we should note that the numerical values of the clustering coefficient C are remarkably larger, than in the case of the data from LiveJournal. The reason for this discrepancy could lay in the fact that while the simulated line graph is uncorrelated, the actual communities and interests grouping LiveJournal users are.

5 Conclusion

As we demonstrated above, the LiveJournal social network is scale-free, clustered and assortative. This makes it similar to the line graph, constructed on a scale-free network. Additionally, this similarity captures also the jagged character of the clustering coefficient dependence $C(k)$ on the node degree k . This similarity suggests, that a line graph, constructed on a scale-free network, is a fair representation of a realistic social network. This is the main goal of this paper.

Aside from suggesting a natural mechanism for the social network construction, a direct application of this result appears, if we are interested in a simulation of the process of spread of information, as alerts or gossips, in a community. For a large network, the direct simulation of the state of each particular node can be burdensome and memory consuming. Instead, we can consider a hypothesis that within the cliques, the information is shared almost immediately, when compared with the time of its transmission between the cliques. Such

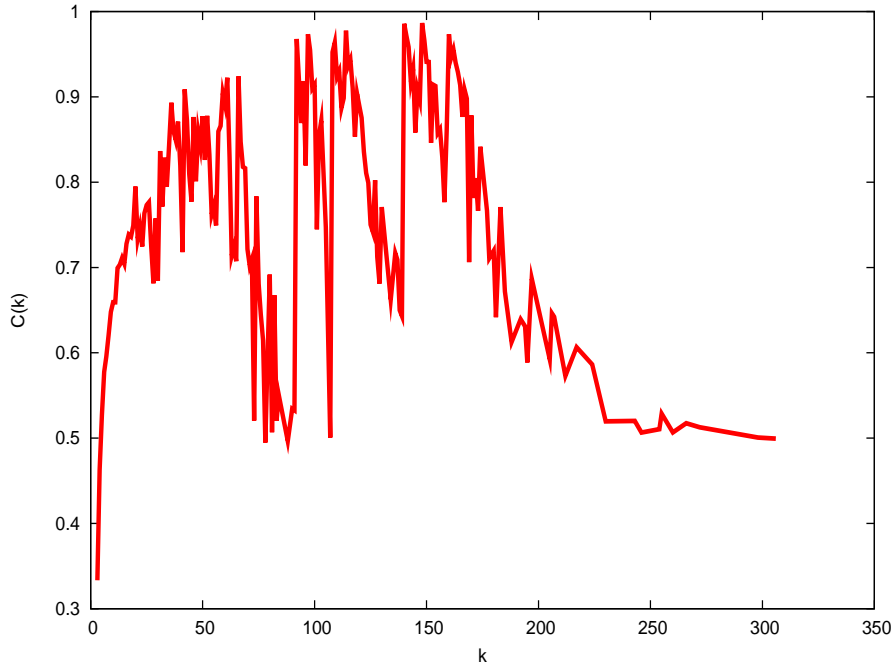


Figure 4: The clustering coefficient $C(k)$ for a line graph constructed on an uncorrelated scale-free network.

mechanism has been suggested by a number of social science researchers. In fact, it is implied by the Granovetter's groundbreaking "strength of the weak ties" [1]. If this is the case, it is possible to simulate the process on a much smaller network, where nodes represent cliques.

Concluding, topological arguments are presented that real social networks, where nodes represent agents, can be modeled as line graphs constructed on initial networks, where nodes represent families, school classes, groups of friends who meet everyday, teams in firms etc. Modeling the spread of information, we can work on the initial networks, which are clearly smaller and more simple.

Acknowledgements

The research is partially supported within the FP7 project SOCIONICAL, No. 231288.

References

- [1] M. S. Granovetter, *The strength of weak ties*, Amer. J. of Sociology 78 (1973) 1360.
- [2] J. Scott, *Social Network Analysis. A Handbook*, Sage Publ., London 1991.

- [3] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge UP, Cambridge 1994.
- [4] L. Freeman, *The Development of Social Network Analysis*, Empirical Press, Vancouver 2006.
- [5] www.insna.org/
- [6] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature 393 (1998) 409.
- [7] A.-L. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*, Plume Books, New York 2003.
- [8] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford UP, Oxford 2003.
- [9] *Handbook of Graphs and Networks: From the Genome to the Internet*, Eds. S. Bornholdt and H. G. Schuster, Wiley-VCH, Berlin 2003.
- [10] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge UP, Cambridge 2004.
- [11] M. E. J. Newman, *Networks. An Introduction*, Oxford UP, Oxford 2010.
- [12] H. Whitney, *Congruent graphs and the connectivity of graphs*, Amer. J. Math. 54 (1932) 150.
- [13] T. S. Evans and R. Lambiotte, *Line graphs, link partitions and overlapping communities*, Phys. Rev. E 80 (2009) 016105.
- [14] M. Morris and M. Kretzschmar, *Concurrent partnerships and transmission dynamics in networks*, Social Networks 17 (1995) 299.
- [15] F. Liljeros, C. R. Edling and L. A. Nunes Amaral, *Sexual networks: implications for the transmission of sexually transmitted infections*, Microbes and Infection 5 (2003) 189.
- [16] J. C. Nacher, N. Ueda, T. Yamada, M. Kanehisa and T. Akutsu, *Clustering under the line graph transformation: application to reaction network*, BMC Bioinformatics 24 (2004) 207.
- [17] J. C. Nacher, T. Yamada, S. Goto, M. Kanehisa and T. Akutsu, *Two complementary representations of a scale-free network*, Physica A 349 (2005) 349.
- [18] S. Zhang, H.-W. Liu, X.-M. Ning and X.-S. Zhang, *A hybrid graph-theoretic method for mining overlapping functional modules in large sparse protein interaction networks*, Int. J. of Data Mining and Bioinformatics, 3 (2009) 68.
- [19] D. Ucar, S. Parthasarathy, S. Asur and C. Wang, *Effective pre-processing strategies for functional clustering of a protein-protein interactions network*, Bioinformatics and Bioengineering, Proc. of the Fifth Symp. on Bioinformatics and Bioengineering, (2005) 129.

- [20] A. Apolloni, K. Channakeshava, L. Durbeck, M. Khan, C. Kuhlman, B. Lewis and S. Swarup, *A study of information diffusion over a realistic social network model*, CSE, vol. 4, pp.675-682, 2009 International Conference on Computational Science and Engineering, 2009.
- [21] M. E. J. Newman and Y. Park, *Why social networks are different from other types of networks*, Phys. Rev. E 68 (2003) 036122.
- [22] A. Mańka-Krasoń, A. Mwijage and K. Kułakowski, *Clustering in random line graphs*, Comp. Phys. Commun. 181 (2010) 118.
- [23] A. Mańka-Krasoń and K. Kułakowski, *Assortativity in random line graphs*, Acta Phys. Pol. B Proc. Suppl. 3 (2010) 259.
- [24] M. E. J. Newman, *Assortative mixing in networks*, Phys. Rev. Lett. 89 (2002) 208701.
- [25] M. E. J. Newman, *Mixing patterns in networks*, Phys. Rev. E 67 (2003) 026126.
- [26] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*, PNAS 99 (2002) 7821.
- [27] R. Guimerá, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas, *Self-similar community structure in a network of human interactions*, Phys. Rev. E 68 (2003) 065103.
- [28] J. R. Tyler, D. M. Wilkinson and B. A. Huberman, *Email as spectroscopy: automated discovery of community structure within organizations*, The Information Society 21 (2005) 143.
- [29] P. S. Bearman, J. Moody and K. Stovel, *Chains of affection: the structure of adolescent romantic and sexual networks*, Amer. J. of Sociology 110 (2004) 44.
- [30] S. Fortunato, *Community detection in graphs*, Phys. Reports 486 (2010) 75.
- [31] V. K. Balakrishnan, *Schaum's Outline of Graph Theory: Including Hundreds of Solved Problems*, McGraw-Hill, New York 1997.
- [32] LiveJournal. Available from: www.livejournal.com.
- [33] Balog, K., G. Mishne, and M.d. Rijke. Why are they excited?: identifying and explaining spikes in blog mood levels. in Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters and Demonstrations table of contents. 2006. Trento, Italy.
- [34] Herring, S.C., et al. Language networks on LiveJournal. in Proceedings of the Fortieth Hawaii International Conference on System Sciences. 2007. Los Alamitos: CA: IEEE Press.
- [35] Hsu, W.H., et al., Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach.

- [36] Kendall, L., Shout Into the Wind, and it Shouts Back: Identity and interactional tensions on LiveJournal. First Monday, September 2007. 12(9).
- [37] Kitsak, M., et al., Identification of influential spreaders in complex networks. Nature Physics, 2010.
- [38] Liben-Nowell, D. and J. Kleinberg., Tracing Information Flow on a Global Scale Using Internet Chain-Letter Data. PNAS, 2008. 105(12): p. 46334638.
- [39] Mishne, G.A., Experiments with mood classification in blog posts. Style2005- 1st Workshop on Stylistic Analysis of Text for Information, 2005.